Mol Genet Genomics (2007) 278:539–553
DOI 10.1007/s00438-007-0270-9

ORIGINAL PAPER

Transcriptome profiling, sequence characterization, and SNP-based chromosomal assignment of the *EXPANSIN* genes in cotton

Chuanfu An · Sukumar Saha · Johnie N. Jenkins · Brian E. Scheffler · Thea A. Wilkins · David M. Stelly

Received: 27 April 2007 / Accepted: 15 June 2007 / Published online: 28 August 2007

Abstract The knowledge of biological significance associated with DNA markers is very limited in cotton. SNPs are potential functional marker to tag genes of biological importance. Plant expansins are a group of extracellular proteins that directly modify the mechanical properties of cell walls, enable turgor-driven cell extension, and likely affect length and quality of cotton fibers. Here, we report the expression profiles of *EXPANSIN* transcripts during fiber elongation and the discovery of SNP markers, assess the SNP characteristics, and localize six *EXPANSIN A* genes to chromosomes. Transcriptome profiling of cotton fiber oligonucleotide microarrays revealed that seven *EXPANSIN* transcripts were differentially expressed when there was parallel polar elongation during morphogenesis at early stage of fiber development, suggesting that major and minor isoforms perform discrete functions during polar elongation and lateral expansion. Ancestral and homoeologous relationships of the six *EXPANSIN A* genes were revealed by phylogenetic grouping and comparison to extant A- and D-genome relatives of contemporary AD-genome cottons. The average rate of SNP per nucleotide was 2.35% (one SNP per 43 bp), with 1.74 and 3.99% occurring in coding and noncoding regions, respectively, in the selected genotypes. An unequal evolutionary rate of the *EXPANSIN A* genes at the subgenome level of tetraploid cotton was recorded. Chromosomal locations for each of the six *EXPANSIN A* genes were established by gene-specific SNP markers. Results revealed a strategy for discovering SNP markers in a polyploidy species like cotton. These markers could be useful to associate candidate genes with the complex fiber traits in MAS.

Keywords *Gossypium* spp. · *EXPANSIN* · Transcriptome profiling · Single-nucleotide polymorphism · Chromosomal assignment

Communicated by R. Hagemann.

Nucleotide sequence data reported are available in GenBank database under the accession numbers EF644199 to EF644335.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-007-0270-9) contains supplementary material, which is available to authorized users.

Disclaimer Mention of trade names or commercial products in this manuscript does not constitute a guarantee or warranty of the product by the US Department of Agriculture and does not imply its approval to the exclusion of other products that may also be suitable.

C. An
Department of Plant and Soil Sciences,
Mississippi State University, Mississippi State,
MS 39762, USA

S. Saha (✉) · J. N. Jenkins
USDA-ARS, Crop Science Research Laboratory,
P.O. Box 5367, Mississippi State, MS 39762, USA
e-mail: ssaha@msa-msstate.ars.usda.gov

B. E. Scheffler
USDA-ARS-CGRU, Stoneville, MS 38776, USA

T. A. Wilkins
Department of Plant Sciences,
University of California, Davis, CA 95616, USA

Present Address:
T. A. Wilkins
Department of Plant and Soil Science,
Texas Tech University, Lubbock, TX 79409, USA

D. M. Stelly
Department of Soil and Crop Sciences,
Texas A&M University, College Station, TX 77843-2474, USA

Introduction

Cotton (*Gossypium* spp.) is the leading natural fiber crop of the world. Approximately 90% of its value resides in the fiber (lint). Although lint production has increased recently, the fiber quality has been declining over the last decade. Fiber quality, determined by length, strength, elongation, and uniformity, is the most important factor in modern spinning technology and profitability. Botanically, fiber is a single-celled trichome developing from individual epidermal cells on the outer integument of cotton ovules. The development of fiber cells undergoes four discrete, yet overlapping stages: differentiation, expansion/primary cell wall (PCW) synthesis, secondary cell wall (SCW) synthesis, and maturation (Wilkins and Jernstedt 1999; Wilkins and Arpat 2005). So far, many genes involved in cotton fiber development have been isolated and characterized (Arpat et al. 2004). There is strong interest from biological and economical standpoints to identify genes for which expression closely parallels the rate of cotton fiber expansion and elongation, since these will likely include genes that impart major influences on cell wall development, fiber quality, and many other plant attributes. Most of the economically important fiber traits, including elongation, are controlled by quantitative trait loci (QTLs). Knowledge of functional genes underlying fiber quality QTLs is very limited in cotton.

Expansins are a large family of extracellular proteins that loosen the components of rigid plant cell walls and thereby allow cell expansion (Darley et al. 2001; Li et al. 2002; Sampedro and Cosgrove 2005). It had been found that the regulation of cell wall extensibility during cell expansion was controlled, in part, by different expression of *EXPANSIN* genes in tomato (Vogler et al. 2003) and cotton (Arpat et al. 2004). These functions suggest that expansins affect economical significant fiber properties such as length and elongation. Following the recommendation of an ad hoc working group on nomenclature of *EXPASIN* gene family, we refer to "α-*expansin*" as *EXPANSIN A* (abbreviated as *EXPA*; Kende et al. 2004). A key research model for cell biogenesis research is the cotton fiber, each of which arises from a single epidermal cell of the ovule integument that commences extensive unipolar expansion on the day of anthesis [0 days post-anthesis (dpa)] and lasts to around 20 dpa (Smart et al. 1998; Wilkins and Jernstedt 1999). Elongation of cotton fibers is highly polar and rapid; growth rates in cultivated species exceed 2 mm/day during peak growth (Wilkins and Jernstedt 1999). Expression of several *EXPANSIN* genes parallels fiber elongation (Shimizu et al. 1997; Orford and Timmis 1998; Ruan et al. 2001; Harmer et al. 2002; Ji et al. 2003). Comprehensive analyses of the cotton fiber transcriptome showed that *GhEXPA1* (AF043284) is one of the top 15% most highly expressed genes in *G. arboreum* L. cv. AKA8401 (Arpat et al. 2004). Given the inferred importance of *EXPANSIN* genes in cotton fiber elongation and the presence of multigene family of *EXPANSIN* in other plant species (Li et al. 2002; Sampedro and Cosgrove 2005), it is reasonable to hypothesize that some *EXPANSIN* genes could be involved in cotton fiber development and that more of these are yet to be discovered.

Molecular markers used in cotton genome mapping and genetic diversity analysis have evolved from hybridization-based RFLPs (Reinisch et al. 1994; Shappley et al. 1998; Rong et al. 2004) to PCR-based markers such as RAPDs (Kohel et al. 2001), AFLPs (Abdalla et al. 2001; Mei et al. 2004), and microsatellites (Zhang et al. 2002; Han et al. 2004, 2006; Park et al. 2005; Frelichowski et al. 2006). For these markers, relatively low levels of intraspecific polymorphism and limited association with candidate genes have hampered integrated genetic mapping and important candidate gene mapping. DNA markers specific to candidate genes will help in the association of biological important genes with complex fiber QTLs. Single-nucleotide polymorphism (SNP), including single-base changes or indels (insertion or deletion) at specific nucleotide positions, has been shown to be the most abundant class of DNA polymorphism in many organisms (Kwok et al. 1996; Wang et al. 1998; Brookes 1999; Cho et al. 1999). SNP variation analysis and SNP marker development from candidate genes could provide valuable information regarding their evolution and effects on complex traits. The anticipated value of SNPs for analysis of candidate gene evolution and their effects on complex traits have stimulated large scale SNP characterization and marker mapping in rice (Feltus et al. 2004), wheat (Mochida et al. 2003; Somers et al. 2003; Zhang et al. 2003; Caldwell et al. 2004), maize (Ching et al. 2002; Batley et al. 2003), soybean (Zhu et al. 2003; Kim et al. 2005), and barley (Kanazin et al. 2002; Bundock et al. 2003; Bundock and Henry 2004). Most cotton sequence variation analyses have been confined to a single gene family or DNA fragments for phylogenetic analysis (Small et al. 1998; Small and Wendel 2000; Cronn et al. 2002; Alvarez et al. 2005). Candidate gene-based association mapping using SNP markers has emerged as a powerful tool to determine the role of genes in complex traits (Glazier et al. 2002). Despite the importance of SNPs in studies on human diseases, few SNP analyses have been carried out in plants, especially in polyploid crops compared to other types of markers (Kanazin et al. 2002; Batley et al. 2003; Neale and Savolainen 2004). The research on SNP analysis in cotton is almost nil due to the large genome size (Grover et al. 2004), tetraploid nature, the presence of high repetitive DNA content (Zhao et al. 1998), and paucity of information on genomic sequences (Chee et al. 2004). Sequence-tagged sites (STS) sequencing

results indicated that the rate of variation per nucleotide was 0.35% between *G. hirsutum* and *G. barbadense*, and the variation per nucleotide was 0.14 and 0.37% within these two species, respectively (Rong et al. 2004).

To further explore the expression patterns and the function of *EXPANSIN* isoforms, we examined the expression profiling of *EXPANSIN* related transcriptome using the microarray technique (Arpat et al. 2004). Here, we apply SNP analysis and marker-based gene chromosomal assignment to the six well-characterized cotton *EXPANSIN A* genes reported by Harmer et al. (2002). The results functionally discriminate isoforms of the *EXPANSIN* gene family and provide useful strategy for distinguishing homoeologous sequences in allotetraploid cotton SNP discovery. The chromosomal locations and SNP markers derived from the six *EXPANSIN A* genes would be useful in integrated genetic mapping and fiber quality-related QTLs analysis in cotton.

Materials and methods

Plant materials, DNA and RNA isolation

Two populations of TM-1 (*G. hirsutum* L.) were grown as biological pools under a 30/21°C day/night temperature regime in a greenhouse at UC Davis. Flowers were tagged on the day of anthesis (0 dpa) to collect developing bolls from the first fruiting positions at 5, 8, 10, 14, 17, 21, and 24 dpa between 8:00 a.m. and 10:00 a.m. to eliminate diurnal effects. In order to minimize sampling variation, the harvesting scheme devised by Arpat et al. (2004) was used and materials stored at −80°C. Fibers were separated from frozen ovules/locules in liquid nitrogen using a mortar and pestle (Smart et al. 1998) to generate 2–3 pools representing 8–18 different bolls. Two total RNA samples were prepared independently from ~0.5 g of fiber following a modification of the hot borate method (Wilkins and Smart 1996) as described (http://www.cfgc.ucdavis.edu/).

Four tetraploid species, TM-1, HS46 and MARCABUCAG8US-1-88 (MAR; *G. hirsutum* L., $[AD]_1$), 3–79 (*G. barbadense* L., $[AD]_2$), *G. tomentosum* Nuttall ex Seemann ($[AD]_3$), *G. mustelinum* Miers ex Watt ($[AD]_4$), and two diploid genome species, *G. arboreum* L. (A_2) and *G. raimondii* Ulbrich (D_5), were used for PCR amplification and SNP marker identification. TM-1 and 3–79 served as genetic standards for *G. hirsutum* and *G. barbadense*, respectively. HS46 and MAR are the two parents of recombinant inbred lines developed for mapping projects (Shappley et al. 1998; Ulloa et al. 2005). *G. tomentosum* and *G. mustelinum* are two wild cotton species. *G. arboreum* and *G. raimondii* are extant relatives of species that donated the A and D genomes of the original AD allotetraploid that gave rise to the modern 52-chromosome *Gossypium* species (Brubaker et al. 1999; Wendel and Cronn 2003). Two kinds of genetic stocks were used for chromosomal assignment of *EXPANSIN A* genes by deletion analysis: (1) Quasi-isogenic hypoaneuploid interspecific F_1 hybrid chromosome substitution stocks, each involving chromosomally identified primary monosomy, monotelodisomy, or tertiary monosomy (Liu et al. 2000; Ulloa et al. 2005), and (2) quasi-isogenic euploid CS-B lines, i.e. BC_5S_1-derived interspecific backcrossed chromosome substitution lines of 3–79 in TM-1 (Stelly et al. 2005; Jenkins et al. 2006; Saha et al. 2006a, b). The primary monosomic plants ($2n = 51$) lacked an entire chromosome of the normal *G. hirsutum* complement, whereas the monotelodisomic plants ($2n = 52$) lacked most or all of just one *G. hirsutum* chromosome arm. Tertiary monosomics were deficient in one of the two reciprocally translocated *G. hirsutum* chromosomes described in detail by Brown et al. (1981) and Menzel et al. (1985). Thus, each tertiary monosomic plant lacked the centric segment from one *G. hirsutum* chromosome and the distal acentric segment from the other chromosome involved in the translocation. We used two sets of monosomic and monotelodisomic F_1 interspecific hybrids, from *G. hirsutum* aneuploids crossed with 3–79 (*G. barbadense*) versus *G. tomentosum*; and one set of tertiary monosomic hybrids, from crosses with *G. tomentosum* only. DNA samples of the diploid species (*G. arboreum* and *G. raimondii*) were kindly provided by Dr. John Yu (USDA-ARS, Crop Germplasm Research Unit, College Station, TX, USA). Genomic DNA of the different genotypes and species were isolated from young leaves of individual plants using a DNeasy Plant Maxi Kit (Qiagen Inc., Valencia, CA, USA).

Microarray expression analysis

Cotton fiber 70-mer cDNA microarrays, based on the cotton fiber EST database, were fabricated by the UC Davis Cotton Functional Genomics Center (Arpat et al. 2004; Table 1). Microarrays were printed by spotting oligonucleotides (40 μM) suspended in 50% DMSO on Corning GAPSII slides in duplication, arranged in 23 × 23 subarrays using the OmniGrid arrayer (Genomic Solutions, Ann Arbor, MI, USA) equipped with 16 (4 × 4) MicroQuill pins (Majer Precision Engineering, Tempe, AZ, USA). Slide post-processing utilized UV exposure to chemically cross-link oligomers according to the slide manufacturer's instructions. Fluorescently tagged amino-allyl-labeled cDNA probes were generated by reverse transcription of cotton fiber total RNA (30 μg), which was mixed with reference mRNA spike mix (10, 50, 200, 400, 1000, and 2000 pg Alien Spike mRNAs 1 through 6, respectively, Stratagene,

Table 1 List of seven cotton fiber *EXPANSIN* transcripts and related microarray oligonucleotide probes from the *Ga* (*G. arboreum*) cotton fiber dbEST (http://www.cfgc.ucdavis.edu)

Probe database reference	EST name
Cotton12_16168_001[a]	GA_Ed0103C11f
Cotton12_12779_001	GA_Eb0009P18
GA_Ed0108H10f	GA_Ed0108H10f
gi|908878|gb|108878|GA_Eb0026E18f	GA_Eb0026E18
gi|7502129|gb|AW667749|GA_Ea0010H23	GA_Ea0010H23
gi|8381468|gb|BE054412|GA_Ea0032B18	GA_Ea0032B18
gi|907832|gb|107832|GA_Eb0025K12f	GA_Eb0025K12

[a] Probe generated from EST GA_Ed0103C11f exactly matched EST DQ023525, AF043284, AY189969, and the two *EXPANSIN A* genes [AF512539 (*GhEXPA1*), AF512540 (*GhEXPA2*)] for which SNP analysis was performed in this study

La Jolla, CA, USA), 5 µg poly-T_{21}VQ primers (Operon, Huntsville, AL, USA), 0.2 mM each of dATP, dCTP, and dGTP, and a 2:3 ratio of 0.2 mM amino-allyl-dUTP/dUTP mix (Sigma, St Louis, MO, USA) and SuperScript II (Invitrogen, Carlsbad, CA, USA) in a total volume of 40 µl and incubated at 42°C for 4 h. Aminoallyl-labeled cDNA was purified using the QiaQuick PCR purification kit (Qiagen Inc., Valencia, CA, USA), and labeled with Cy3 or Cy5 monofunctional dyes (Amersham, Piscataway, NJ, USA) according to the manufacturer's instructions. Labeled cDNAs were quality controlled by measuring absorbance at 260, 280, 555, and 650 nm. Concentrations of cDNAs were calculated for each probe, and the frequency of incorporation (FOI) for dye molecules was expressed as the number of dye molecules per 100 bp. Probes having an FOI less than 20 were discarded. Printed slides were pre-hybridized for 45 min at 42°C in a sealed coplin jar containing a 50 ml solution of 5 × SSC, 0.1% SDS, and 0.1 µg/µl BSA, and then washed twice for 10 s each in filtered ddH_2O and dried with filtered compressed air. The final hybridization solution (50 µl) contained amino-labeled cDNA in 30% formamide, 5 × SSC, 0.1% SDS, and 0.1 µg/µl sheared salmon sperm DNA and was applied to the slide surface by using mSeries LifterSlips (Erie Scientific, Portsmouth, NH, USA). After incubating at 42°C for 16–20 h in a humidified hybridization chamber (Genetix, Boston, MA, USA), the LifterSlips were removed by dipping slides in 2 × SSC at 42°C. Slides were washed sequentially at 42°C in 2 × SSC, 0.2% SDS for 5 min, 0.5 × SSC for 3 min, four times in 0.1 × SSC for 1 min per wash, and 0.05 × SSC for 10 s, followed by a dip in ddH_2O and drying with filtered compressed air. Slides were scanned at a 10 µm resolution using an Affymetrix Array Scanner 428 (Affymetrix, Santa Clara, CA, USA) and signal intensities were quantified with ImaGene 4.2 (BioDiscovery, El Segundo, CA, USA). The non-linear approach was employed for microarray data analysis by GNU statistical computing language R version 2.0.1. A bi-directional double-loop experimental design was adopted for microarray analysis (Kerr and Churchill 2001; Glonek and Solomon 2004) to analyze all possible significant changes in gene expression between any two developmental time points. The results were interpreted using the 5 dpa time point as the common reference. Control experiments from the same population of plants indicated no significant biological or technical variability (Arpat et al. 2004). Comparison of log 2 signal intensity ratios derived from direct and indirect paths for each developmental time point were well within acceptable ranges indicating good experimental reliability (König et al. 2004).

PCR primer design, amplification, cloning and sequencing

A total of 13 primer pairs were designed for six cotton *EXPANSIN A* genes (Table 2) using Primer 3 software (http://www.frodo.wi.mit.edu/). For each gene, there was at least one primer pair designed to amplify 400–800 bp for sequencing. Gene-specificity of each primer was tested using BLASTN against cotton genomic sequences in GenBank. *Pfu* polymerase (Stratagene, La Jolla, CA, USA) was used for PCR amplification according to the manufacturer's protocol on a PTC-225 Peltier Thermal Cycler (MJ Research Inc, Waltham, MA, USA). The PCR products were excised from agarose gels following electrophoresis, purified using QIAEX II gel extraction kit (Qiagen Inc, Valencia, CA, USA), and cloned into the pCR4-TOPO vector (Invitrogen, Carlsbad, CA, USA) after adding the 3′-A to the purified DNA fragment according to the manufacturer's protocol. Plasmid DNA isolated from kanamycin-resistant colonies was bi-directionally sequenced with ABI Prism BigDye Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems, Foster City, CA, USA) on an ABI automated sequencer. PCR products were initially cloned to enable distinction among related loci and alleles expected in allotetraploid cotton, including pairs of homoeologous At and Dt loci, and any additional duplications or heterozygosity. In this study, 12 recombinant colonies were sequenced for each *EXPANSIN A* gene generated from each genotype to avoid possible complications due to PCR recombination (Cronn et al. 2002). Forward and reverse matched sequences from at least three clones were used to determine the sequence for each duplicated copy (Cedroni et al. 2003; Rong et al. 2004).

Sequence analysis and SNP primer design

Alignment of diverse sequences was conducted using ClustalX (Thompson et al. 1997), and phylogenetic relationships were analyzed by the neighbor joining (NJ) method

Table 2 PCR primers designed from six *EXPANSIN A* cotton fiber genes

Gene	NCBI accession[a]	Primer	Primer sequence (5′→3′)
GhEXPA1	AF512539	GhEXPA1	F: CCCCACGAGAACACTTTGAT
			R: CTAATGGCACTTGCTTGCCT
GhEXPA2	AF512540	GhEXPA2-1	F: GTCAGCCAATTGTTTGAGCTA
			R: TAGATAAAGCATAGTTAGGGG
		GhEXPA2-2	F: ACAGCCACCAACTTTTGTCCC
			R: AGTTTGTCCGAATTGCCAAC
GhEXPA3	AF512541	GhEXPA3-1	F: GTATGCTTTTTGGTATGCAG
			R: GTGTGTCGGTGGAAAATG
		GhEXPA3-2	F: TTTGACAATGGCTTGAGCTG
			R: TCCGTTACTGGTTGTGACGA
GhEXPA4	AF512542	GhEXPA4-1	F: TACGCCCGATATTCAACACA
			R: CGTTTGCGCACTTAATCTCA
		GhEXPA4-2	F: TGAGATTAAGTGCGCAAACG
			R: GCCAGTTTTGACCCCAGTTA
		GhEXPA4-3	F: TGAAGGTGAAGGGAACCAAC
			R: CCCAACCCCCATTTTTACTT
GhEXPA5	AF512543	GhEXPA5-1	F: GTGCCCACCCAATAATTAAA
			R: ATGTTAATCGTACCTCCGAT
		GhEXPA5-2	F: TGATTTTCGAAGGGTGCCAT
			R: TATTGCATGCTCCCAAACAC
GhEXPA6	AF512544	GhEXPA6-1	F: CTGTTGTTTGTTCGCAGGAA
			R: GAAGCAGCAAAAGGCAAAAC
		GhEXPA6-2	F: TGGCCACTCCTACTTCAACC
			R: GACCCGAAAGTCCCACTACA
		GhEXPA6-3	F: GTTTTGCCTTTTGCTGCTTC
			R: GAGAGGCTTTGTCCGTTGAG

In this report we used the standard nomenclature of the *EXPANSIN* genes (Kende et al. 2004)

[a] The NCBI accession numbers represent the original genes isolated from *G. hirsutum* L. cv Siokra 1–4 (Harmer et al. 2002)

using MEGA version 3.1 (Kumar et al. 2004). The putative assignment of a sequence to a particular locus or subgenome was based on phylogenetic analysis and the relationship to diploid ancestral species (A_2 and D_5) of the tetraploid cotton. DnaSP 4.0 software (Rozas et al. 2003) was used to identify SNPs based on a comparative alignment of sequences at a putative locus, and estimate of nucleotide and haplotype diversity. Final average nucleotide diversities (π) were calculated from all pairwise comparisons (Tajima 1983; Nei 1987). Interspecies SNP primers were designed based on single nucleotide differences in the sequences at a putative locus between TM-1 and 3–79 or TM-1 and *G. tomentosum* for chromosomal assignment of SNP markers. Primers were selected to anneal immediately upstream or downstream of the SNP site as the forward or reverse primer, respectively, so that the polymorphism could be detected by one-base extension technology with the ABI Prism SNaPshot™ multiplex kit. Selected primers were evaluated by Primer 3 (http://www.frodo.wi.mit.edu/) using criteria of a primer length of 18–26 nt (20 nt as the optimum), an optimum annealing temperature of 50°C, and a 40–60% GC content.

SNP genotyping

The ABI Prism SNaPshot™ multiplex kit and an ABI 3100 capillary electrophoresis system were used for screening SNP markers following a slight modification of the manufacturer's protocol. PCR products amplified using *Pfu* polymerase were purified by incubation with SAP and *Exo* I (2 units of SAP and 4 units of *Exo* I in a 20 µl PCR reaction volume) at 37°C for 1 h and then at 75°C for 15 min. The thermal cycle reaction mixture (7 µl reaction system in 384-well plates) contained 1.5 µl of SNaPshot multiplex ready reaction mix, 0.5 µl of purified PCR product, 0.2 µl of SNP primer (10 µM), and 4.3 µl of distilled water. The thermal cycle reaction was carried out for 25 cycles at 96°C for 10 s, 50°C for 5 s, and 60°C for 30 s. After treatment with SAP (1 unit) at 37°C for 1 h, and followed by incubating the reactions at 75°C for 15 min, 1 µl of 10-fold diluted SNaPshot product was mixed with 0.2 µl of GeneScan-120 LIZ size standard, and 8.8 µl of Hi-Di formamide, denatured at 95°C for 5 min, and then loaded onto the ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) in SNaShot mode for SNP marker analysis.

Chromosomal assignment

The deletion analysis method was used for identifying chromosomal locations of six *EXPANSIN A* genes (Liu et al. 2000; Ulloa et al. 2005). All of the aneuploid chromosome substitution F_1 lines, except the particular aneuploid line missing a specific chromosome or chromosome arm, would expectedly show two alleles originating from both parents, similar to an F_1 heterozygous locus. The absence of the TM-1 (*G. hirsutum*) allele in any one of the aneuploid F_1 plant indicated that the missing chromosome or chromosome arm was the most likely location of the gene of interest. In euploid CS-B stocks, the absence of the TM-1 allele but presence of the 3–79 allele suggested that the gene was probably located in the substituted chromosome or chromosome arm.

Results

Fiber transcriptome expression profiles of *EXPANSIN* genes

A search of >38,000 cotton (*G. arboreum*) fiber ESTs revealed the presence of seven different *EXPANSIN* transcripts (Table 1). *In silico* expression analysis of fiber ESTs revealed that cell wall-related genes, including two *EXPANSIN* genes (two Ga_Ea clones in Table 1), represent, by far, the most abundant gene transcripts during this stage of development (Arpat et al. 2004). Developmental transcriptome profiles of all *EXPANSIN* genes were extracted from microarray data spanning 5–24 dpa cotton fibers. Changes in the temporal expression of all seven *EXPANSIN* genes in developing fibers of the tetraploid TM-1, relative to 5 dpa are shown in Fig. 1. The expression profiles were an average of the At- and Dt-homoeologs. In general, developmental regulation of *EXPANSIN* gene expression (Fig. 1) closely parallels that of the growth rate during the period of rapid polar elongation (Wilkins and Arpat 2005). The *EXPANSIN* genes were differentially regulated and could be classified as major and minor isoforms. Temporal patterns revealed two major expression trends. After 17 dpa, expression decreased dramatically for the four major *EXPANSIN* isoforms that are most abundant during rapid fiber elongation, specifically GA_Eb0026E18, GA_Ea0032B18, GA_Ed0103C11f, and GA_Eb0025K12 (Fig. 1). By contrast, expression remained relatively high after 21 dpa and well into the stage of secondary cell wall synthesis of relatively minor isoforms, GA_Eb0009P18, GA_Ed0108H10f, and GA_Ea0010H23. Thus, the fiber *EXPANSIN* genes were differentially expressed and followed discrete temporal patterns. Their expression overlapped the termination of elongation and the onset of secondary cell wall biogenesis.

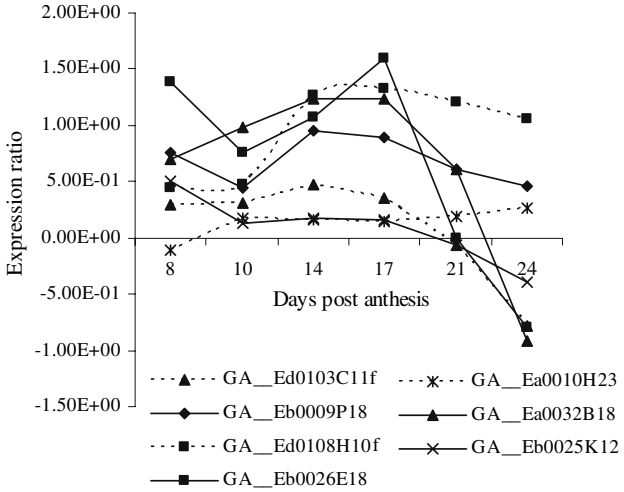

Fig. 1 Developmentally regulated cotton fiber. *EXPANSIN* genes are differentially expressed during the period of primary expansion and rapid polar elongation from anthesis (0 dpa) to 24 dpa fibers. The expression of major isoforms of *EXPANSIN* (GA_Eb0026E18, GA_Ea0032B18, GA_Ed0103C11f, and GA_Eb0025K12) during polar elongation decreased dramatically after 17 dpa, while the expression of relatively minor isoforms (GA_Eb0009P18, GA_Ed0108H10f, and GA_Ea0010H23) remained relatively high after 21 dpa and well into the stage of secondary cell wall synthesis. Peak growth occurred around 14–17 dpa. Log 2 expression ratios were obtained from microarray data and expressed relative to the 5 dpa time point

Sequence characterization of six *EXPANSIN A* genes

Sequence data were obtained from cloned PCR fragments amplified from *EXPANSIN A* genes in selected cotton genotypes using gene-specific primer pairs (Table 2). First, the identities of all the sequences in our results were confirmed to the respective original gene by BLASTN. Phylogenetic analyses were then used to discriminate homoeologous gene sequences based on the relationship with the ancestrally related extant genomes A_2 and D_5. Within each genome cluster, sequences from the available genotypes were classified to one putative locus based on the clustering results of the phylogram. No additional duplicated loci were found within either of the At- or Dt-genomes for any of the six *EXPANSIN A* genes. Any locus-specific sequence difference identified between the selected genotypes was considered as an SNP. Genome assignments and phylogenetic analyses of all PCR-amplified gene fragments are summarized in supplementary Fig. 1.

In total, 52.9 kb of DNA sequences (33.7 and 19.2kb from coding and noncoding regions, respectively) were acquired from both genomes after eliminating overlapping sequences (Table 3). The available sequence from A- and D-genomes was 31.7 and 21.2 kb, respectively. BLASTN analysis showed that amplified gene segments, ranging in size from 340 to 607 bp, had 98–100% nucleotide identity to the target gene. A total of 222 SNPs, including 120 single-base changes and 102 indels, were identified in 134

Table 3 Distribution and types of SNPs in PCR amplicons of six cotton *EXPANSIN A* genes

Fragment	Genome	No. of aviailable sequences	Haplotypes[a]		Transitions		Transversions				Indels				Coding region		Total	
			H[b]	H_d[c]	A/G	T/C	A/T	G/C	A/C	G/T	A	C	G	T	Length (bp)	No. of SNPs (non-synonymous changes)	Length (bp)	No. of SNPs
EXPA1	A	4	3	0.833 (± 0.222	3	1	0	0	0	1	1	1	0	1	392	4(3)	607	8
	D	7	5	0.857 ± 0.137	4	4	0	2	0	2	0	0	0	1	392	7(6)	607	13
EXPA2-1	A	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	D	7	3	0.667 ± 0.160	1	0	0	0	1	0	0	0	0	0	358	2(1)	474	2
EXPA2-2	A	4	4	1.000 ± 0.177	4	1	0	1	0	1	1	1	0	0	435	6(3)	511	9
	D	7	5	0.857 ± 0.137	4	2	0	2	0	2	1	0	0	0	435	10(5)[d]	511	11
EXPA3-1	A	7	3	0.524 ± 0.209	1	0	0	1	1	0	1	2	0	0	416	2(1)	595	6
	D	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
EXPA3-2	A	6	3	0.600 ± 0.210	1	0	0	0	1	0	0	0	0	0	444	2(2)	531	2
	D	5	3	0.700 ± 0.218	1	1	0	0	1	0	0	0	0	0	444	2(1)	527	3
EXPA4-1	A	7	4	0.810 ± 0.130	0	2	0	0	1	0	0	0	0	1	293	3(1)	552	4
	D	1	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
EXPA4-2	A	6	6	1.000 ± 0.096	1	2	2	2	0	0	0	0	0	0	383	4(3)	538	7
	D	6	5	0.933 ± 0.122	0	3	2	0	1	1	0	0	0	0	383	2(0)	538	7
EXPA4-3	A	3	N/A	N/A	0	0	1	0	0	0	0	0	0	0	204	0(0)	340	1
	D	7	5	0.905 ± 0.103	1	1	3	1	0	1	0	0	0	0	204	3(2)	340	7
EXPA5-1	A	7	4	0.714 ± 0.180	3[e]	0	0	1	1	1	1	0	0	1	310	6(5)	420	8
	D	5	3	0.700 ± 0.218	1	2	0	0	0	0	11	12	7	10	310	37(1)[f]	447	43
EXPA5-2	A	6	4	0.867 ± 0.129	0	0	1	0	2	0	1	0	0	0	313	2(2)	509	4
	D	3	2	0.667 ± 0.314	0	1	1	0	0	0	1	0	0	0	313	2(2)	510	3
EXPA6-1	A	5	4	0.900 ± 0.161	3	4	0	2	0	1	0	0	1	0	440	8(4)	572	11
	D	7	5	0.857 ± 0.137	4	2	1	0	0	1	0	0	0	0	440	7(5)	572	8
EXPA6-2	A	5	3	0.700 ± 0.218	1	0	0	0	0	1	0	0	0	0	265	2(2)	417	2
	D	6	2	0.333 ± .215	1	1	0	0	0	0	0	2	0	1	265	2(2)	420	5
EXPA6-3	A	7	4	0.714 ± 0.181	2	3	0	0	1	2	12	2	8	19	204	2(2)	591	49
	D	7	4	0.810 ± 0.130	2	1	2	0	0	2	1	0	0	1	204	4(3)	550	9

[a] When the number of SNPs in the amplified fragment was less than two, haplotype analysis was not conducted. Sites with alignment gaps were not considered for haplotypte analysis

[b] H means haplotype number

[c] H_d refers to haplotype diversity

[d] *G. tomentosum* and *G. raimondii* have one gap in the coding region, which was not considered in the amino acid change analysis

[e] There was one triallelic SNP (A/C/G) site. Here, we recorded it at the A/G transitions position

[f] Thirty five gaps appeared between *G. tomentosum* and *G. raimondii* were not consider in the amino acid change analysis

amplicons (Table 3). Transitions accounted for 69 (57.5%) and transversions for 51 (42.5%) of the total 120 single-base changes. The ratio of "A/G" to "T/C" transitions was 1.23:1, with no significant difference between the four types of transversions. Analysis of indel sequences indicated a bias toward "A" and "T" nucleotides, which is similar to the "A" nucleotide bias in maize (Batley et al. 2003). The average rate of SNPs per nucleotide was 2.35%, with 1.74 and 3.99% occurring in the coding regions and noncoding regions, respectively, in selected genotypes. Based on the average rate of SNP per nucleotide, a higher nucleotide change rate was discovered in the D-genome (2.90%) than in the A-genome (1.98%). Furthermore, an uneven distribution of SNPs was observed among the six genes, suggesting that the occurrence of SNP varies among the six *EXPANSIN A* genes. Amplicons of EXPA2-2 and EXPA5-1 in *G. tomentosum* and *G. raimondii* had a 46- and 35-nucleotide deletion, respectively, thereby contributing to a higher rate of nucleotide polymorphism in the Dt-genome of tetraploid species. One putative triallelic SNP site ("A/C/G") was discovered in the A-genome amplicon of EXPA5-1, similar to triallelic SNP loci observed in soybean ESTs (Van et al. 2005). The simple sequence repeat ("TA") motif is present in the intron of the EXPA5-1 amplicon, as previously reported (Kumar et al. 2006). The 119 cSNPs included 83 single-base changes and 36 indels in the 33.7 kb of coding sequence. Among the single-base changes, 26, 31, and 26 were detected in the first, second, and third codon positions, respectively. Results revealed that a total of 56 out of the 119 cSNPs (47%) were nonsynonymous changes, provided that the aforementioned fragment deletions of EXPA2-2 and EXPA5-1 in *G. tomentosum* and *G. raimondii* were not considered.

Detailed results of putative haplotype organization are included in the supplementary Tables 1–22. The number of SNP haplotypes present in each group was determined when the available sequence number exceeded two and more than one SNP variable site presented in the aligned sequences. Results showed that the haplotype number ranged from two to six out of the maximum seven available genotype sequences, with an estimated haplotype diversity that varied between 0.33 and 1 (Table 3). The relatively high haplotype number indicated the distinct and diverse SNP characters among the selected cotton species, especially at the interspecies level, which represents valuable sources of germplasm diversity for upland cotton improvement.

Nucleotide diversity

Nucleotide diversity (π) across all possible comparisons among the eight genotypes was measured using available sequences from A- or D-genomes. Because only two sequences were compared at a time, and no segregating populations were considered, we only report the π value for the nucleotide diversity assay, which measures the average number of nucleotide differences per site between two sequences (Nei 1987). The mean value of π in the A- and D-genome is summarized in Table 4. Results showed that nucleotide diversity was lowest among the three *G. hirsutum* lines (TM-1, HS46, and MAR), revealing that polymorphisms among the six *EXPANSIN A* genes were much higher at the interspecific level than intraspecific level in *Gossypium* species. The lowest nucleotide diversity was found between HS46 and MAR in both At- and Dt-genomes, observed as 0.0207 and 0.1018, respectively. Nucleotide diversities (π values) among the three *G. hirsutum* lines in the Dt-genome were significantly higher than those in the At-genome, indicating that loci in the Dt-genome exhibit a faster evolutionary rate among the six *EXPANSIN A* genes compared to the At-genome in tetraploid cotton.

The independent and incongruent evolution of the two subgenomes (At and Dt) was also revealed by the different

Table 4 The nucleotide diversity (π values $\times 10^{-2}$) among six PCR-amplified cotton *EXPANSIN A* gene fragments in A- and D-genomes

Genotype[a]	*G. arboreum*	TM-1	HS46	MAR	3–79	*G. mustelinum*	*G. tomentosum*
G. raimondii	–	0.5622	0.5457	0.7105	0.6166	0.7731	0.6812
TM-1	0.5137	–	0.1243	0.2167	0.5916	0.6417	0.4699
HS46	0.4759	0.031	–	0.1018	0.5087	0.5491	0.4080
MAR	0.5175	0.071	0.0207	–	0.5994	0.6694	0.4160
3–79	0.6353	0.2964	0.4073	0.3826	–	0.3659	0.3663
G. mustelinum	0.5461	0.3103	0.2363	0.2967	0.3841	–	0.4595
G. tomentosum	0.6689	0.3494	0.3027	0.3029	0.3209	0.392	–

π value, a measure of the average number of nucleotide differences per site between two sequences. The numbers below and above the diagonal represent A- and D-genome sequences pairwise comparisons, respectively

[a] *G. arboreum* and *G. raimondii* were considered related to the diploid ancestral A- and D-genome progenitors of tetraploid species; TM-1 was considered as genetic standard of *G. hirsutum* and 3–79 was includes as a representative sample of *G. barbadense*; HS46 and MAR (abbreviation of MARCABUCAG8US-1-88) were used as two diverse *G. hirsutum* lines. One accession of *G. mustelinum* and *G. tomentosum* were included to represent another two allotetraploid species

phylogenetic topologies detected in the duplicated genes or fragments of the selected tetraploid species. The highest nucleotide diversity in the A-genome was observed between the ancestral A-genome species (*G. arboreum*) and *G. tomentosum* ($\pi = 0.6689$), while in the D-genome, the highest nucleotide diversity was obtained from *G. raimondii* (ancestral D-genome donor) and *G. mustelinum* ($\pi = 0.7731$) (Table 4). When considering all other genotypes, the different phylogenetic patterns of At- and Dt-genomes became even more evident. Analyses of the six *EXPANSIN A* genes suggest that polyploid speciation in cotton was accompanied by an independent subgenome molecular evolution.

Chromosomal assignment

We used single-nucleotide extension with primers shown in Table 5 to genotype 25 gene-specific SNP markers in the six *EXPANSIN A* genes. The chromosomal location of each SNP locus was delimited by deletion analysis with one or more hypoaneuploid or euploid stocks, and the results were compared for SNPs within and across the six *EXPANSIN A* genes. The individual SNP loci were assigned to the long arms of chromosomes 20, 10, 9, 1, and 3 (Table 5). For the four genes (*GhEXPA3*, *GhEXPA4*, *GhEXPA5*, and *GhEXPA6*) and the At-genome locus of gene *GhEXPA2*, represented here by multiple SNPs, chromosomal assignments were concordant among SNPs within a gene. SNP markers Exp1-1_Gbmt_193F, and Exp2-1_Gbmt_378R from gene *GhEXPA1* and Dt-genome locus of gene *GhEXPA2*, respectively, were both localized to the chromosome arm 20Lo, which indicated the chromosome locations of the two genes. Chromosomes 10 and 20 are homoeologous, so it is possible that we detected At- and Dt- homoeologous loci of *GhEXPA2* on the long arms of chromosomes 10 and 20. In

Table 5 Chromosomal locations of six *EXPANSIN A* genes in cotton

Gene	SNP marker[a]	SNP primer sequence (5′→3′)	Chromosome location[b]		
			Aneuploid *G. barbadense*	Aneuploid *G. tomentosum*	Euploid CS-B
GhEXPA1	Exp1-1_Gbmt_193F	GTCCGAATTGCCAACCAGC	20*Lo*	20*Lo*	N/A
GhEXPA2	Exp2-1_Gbmt_378R	CACTTTTCTTCTTTTTGTTCAGT	20*Lo*	20*Lo*	N/A
	Exp2-2_Gbt_58F	CAGCAGGCACTACATTGTAG	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_59R	AGCGATGGCAGGACTATCACA	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_93F	CATCGCTGGCAGTCACTTT	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_108R	AGAGCAATGCTTACCTTAACGG	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_175F	CTGGACATAGGTAGCCATCCTGTT	10*Lo*	10*Lo*	10
	Exp2-2_Gb_182R	GATATAACGTCAGTGTCCATCAAG	N/A	N/A	10
	Exp2-2_Gbt_312F	TGACACCCTGCAAAAGGT	10*Lo*	10*Lo*	10
	Exp2-2_Gbmt_345R	AAACTCAATTCAAATCATCAC	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_415F	ACGATTCCAGCTCGATATTC	10*Lo*	10*Lo*	10
	Exp2-2_Gbmt_422R	CCGAACCGGCATTCTTGC	10*Lo*	10*Lo*	10
	Exp2-2_Gbt_508R	ACAGCCACCAACTTTTGTCC	N/A	10*Lo*	10
GhEXPA3	Exp3-1_Gb_489F	TGATCTCTCTCAGCCTATTTTT	10*Lo*	N/A	10
	Exp3-2_Gb_372F	TCTGTATTGGGCAATGTGTT	N/A	N/A	10
GhEXPA4	Exp4-1_Gbmt_65F	GCCATTATTGAAAAGTGCAG	9*Lo*	9*Lo*	9
	Exp4-1_Gbmt_147R	ATGGTGTTTGAATTTTTTT	9*Lo*	N/A	9
	Exp4-1_Gbm_412F	GCATTCACCACAGCCAAAAA	9*Lo*	N/A	9
	Exp4-2_Gbmt_244F	CAAAATCTTTCCCCTTTTACT	9*Lo*	9*Lo*	9
GhEXPA5	Exp5-1_Gmt_205R	CTACCAACTCAGGTGCGATTAC	N/A	1	N/A
	Exp5-2_Gbt_444F	ATACGTCATTAAATTTTCCC	1*Lo*	N/A	1
GhEXPA6	Exp6-1_Gb_77F	CTCGCGATGGTCTCTGGT	3*Lo*	N/A	N/A
	Exp6-1_Gbm_89F	GGTCTCTGGTGTTCAGGGATAT	3*Lo*	N/A	N/A
	Exp6-1_Gbmt_96R	TTGCATGTGCATTAGTCCAA	3*Lo*	3*Lo*	N/A
	Exp6-1_Gb_156R	CTAAAAGATGGCTTCATTTGAAGC	3*Lo*	N/A	N/A

[a] The nomenclature of the SNP markers followed the order: amplified fragment, polymorphic character, SNP site position in the amplified fragment, and forward or reverse primer. For example, Exp2-2_Gbmt_422R means: this marker is located at position 422 of amplified fragment EXPA2-2; this SNP site is polymorphic between TM-1 and 3–79, *G. mustelinum* or *G. tomentosum*; it is a reverse amplification primer

[b] 20*Lo* means on the long arm of chromosome 20

a few cases where aneuploid analysis could not be confirmed by euploid CS-B lines or visa versa, chromatin losses during backcrossing or other types of cytological abnormalities in the development of these cytogenetic stocks could explain the results, suggesting that these particular cytogenetic stocks warrant further characterization.

Discussion

Cotton fiber elongation is the net result of the complex interplay between cell turgor and cell wall extensibility. The elongation is coupled with the expression of many genes, among which *EXPANSIN* is one of most highly expressed (Arpat et al. 2004). To more fully explore and differentiate among *EXPANSIN* genes, we characterized *EXPANSIN* transcriptome expression profiles during cotton fiber polar elongation, assessed SNPs and developed SNP markers in the six well-characterized cotton *EXPANSIN A* genes and chromosomally assigned each of them. Success of the SNP marker development approach used here indicates a workable strategy to distinguish homoeologous sequences and thus SNP marker development for genes within multigene family. The chromosome localization results will contribute to the comparative map of cotton chromosomes and facilitate research that specifically tests whether any of the six *EXPANSIN A* genes impact fiber quality, e.g., by enabling fiber quality SNP-QTL association analysis.

EXPANSIN transcriptome expression profiling

Our profiling of *EXPANISN* transcripts by quantitative analysis of expression cDNA microarrays was inferably more accurate than previous profiling efforts (Shimizu et al. 1997; Orford and Timmis 1998; Ruan et al. 2001; Harmer et al. 2002; Ji et al. 2003), which relied on visually comparing the brightness of the bands from RT-PCR or RNA gel blotting assays. The microarray-based expression profiles revealed parallels between seven different *EXPANSIN* transcripts and polar elongation during fiber morphogenesis. The time difference between the highest rate of cotton fiber elongation (approximately 10–12 dpa; Wilkins and Arpat 2005) and *EXPANSIN* transcriptome expression peak (around 17 dpa) indicated that the correlation between *EXPANSIN* gene transcripts accumulation and maximum fiber growth rate was limited, as was observed in tomato for *LeExp2* and *LeExp18* gene expression (Caderas et al. 2000). In addition to being developmentally regulated, our results showed that the stage-specific *EXPANSIN* genes are also differentially expressed (Fig. 1). Wilkins and Arpat (2005) proposed that the members of gene families like *EXPANSIN* were temporally regulated, indicating that closely related genes play functional roles in cotton fiber development, e.g., stage-specific expression. Wilkins and Arpat (2005) divided the *EXPANSIN* isoforms into major and minor groups, which were responsible for primarily polar elongation and lateral expansion, respectively. In this case, transcripts GA_Eb0026E18, GA_Ea0032B18, GA_Ed0103C11f, and GA_Eb0025K12 were the major isoforms, and the minor isoforms came from the other three transcripts in the *EXPANSIN* multigene family (Fig. 1).

In cotton, more than 60 *EXPANSIN* genes or gene-like sequences have been deposited to the NCBI database so far. Considering the *EXPANSIN* ESTs in *Ga* (*G. arboreum*) cotton fiber dbEST (http://www.cfgc.ucdavis.edu), it is reasonable to conclude that cotton *EXPANSIN* genes are also a complex superfamily. The *EXPANSIN* transcriptome profiles assay made the study of *EXPANSIN* genes involved in cotton fiber development possible. BLASTN analysis showed that the probe generated from EST GA_Ed0103 C11f exactly matched EST DQ023525, AF043284, AY189969, and the two *EXPANSIN A* genes [AF512539 (*GhEXPA1*), AF512540 (*GhEXPA2*)] for which SNP analysis was performed in this study. So, *GhEXPA1* or *GhEXPA2* may be classified as the major isoforms of *EXPANSIN* and function primarily in polar elongation, which may well account for the bulk of genetic variability associated with major QTLs for fiber quality (Harmer et al. 2002; Wilkins and Arpat 2005). The other probes seem unique to the remaining six transcripts because there is no significant identified match in the database. The gap between transcriptome assay and gene isolation and characterization of *EXPANSIN* in cotton indicated the complexity of this multigene family and furthermore their important role in fiber development. Cloning new gene members from the *EXPANSIN* superfamily is still possible and necessary for studying developing cotton fiber. Considering the tetraploid nature of upland cotton, discerning the expression among all the *EXPANSIN* genes and their homoeologous genes is recommended.

SNP marker discovery in tetraploid cotton

Polyploidy complicates the identification of SNPs by the fact that highly similar sequences exist in different subgenomes (homoeologous sequences) or duplicate copies within a genome (paralogous sequences; Mochida et al. 2003; Somers et al. 2003). So far, two common methodologies have been employed to overcome this barrier. One is locus-specific PCR amplification and the other is an *in silico* approach based on clustering that distinguishes paralogs (Richert et al. 2002; Mochida et al. 2003; Somers et al. 2003; Caldwell et al. 2004). However, the high sequence conservation of homoeologous loci that occurs in both intergenic and genic regions of tetraploid cotton hampered

the design of genome-specific PCR primers for SNP identification (Grover et al. 2004).

Phylogenetic analyses of duplicated low- and single-copy sequences in cotton showed that homoeologs exhibit independent evolution. Cronn et al. (1999) proposed that most genes duplicated during polyploidization in *Gossypium* were expected to exhibit independent evolution in the allopolyploid nucleus. The phylogenetic analysis on six cotton *MYB* transcriptional factors supported this conclusion (Cedroni et al. 2003). Results of this study, which were obtained in 13 amplified fragments from six *EXPANSIN A* genes, also indicated an independent evolution pattern. Here, we identified SNPs based on the theory of independent gene evolution in the At- and Dt-genomes of tetraploid species, combined with an integrated phylogenetic approach that incorporated orthologous sequences of ancestral diploid species represented in the closest living descendants (extant species) of the donor species as a base reference. Putative SNP identification was validated by deletion analysis for chromosomal assignment of the six *EXPANSIN A* genes. The success of this approach was facilitated by the inbred nature of the plant materials (highly inbred lines or doubled haploid line) and use of gene-specific PCR primers, which avoided problems inherent to heterozygous alleles and orthologous sequences. Sequence assembly of *G. arboreum* fiber ESTs generated around 14,000 candidate genes, which offer a vast potential for identifying genes that affect fiber quality (Arpat et al. 2004). The identification of candidate genes is relatively difficult in tetraploid species. Only a very small percent (~5%) of cotton fiber genes have been genetically mapped (Rong et al. 2004) due to the low polymorphism nature. Even more disappointing is the fact that <1% of the ~3,000 fiber genes identified by transcriptome profiling as being developmentally regulated and therefore representing important candidate genes show a lower than average polymorphism when using traditional RFLP and other DNA markers (Alabady and Wilkins, unpublished data). The SNP discovery strategy presented here holds great promise for developing SNP markers from important fiber candidate genes, even though discovery is complicated by polyploidy. The success of this pilot study also indicates that SNP is applicable to large diverse gene superfamilies and eliminates the need to focus only on single- or low-copy genes in cultivated cotton species.

SNP features in cotton

Analysis of DNA sequence diversity among a subset of *EXPANSIN A* genes in diploid and tetraploid cotton revealed a mean SNP frequency of 2.35% (1 SNP per 43 bp of sequence), with 1.74 and 3.99% occurring in the coding regions and noncoding regions, respectively, in selected cotton lines. Depending on the species and the genomic region under investigation, recent studies indicated that SNP frequencies range from as high as 1 SNP in 10–15 bp in some noncoding regions of the human genome (Brookes 1999) to a much lower 1 SNP per 4 kb in some highly conserved regions (Nickerson et al. 1998). In plants, SNP frequency also varies among species and is distributed unevenly across genomes. The average SNP frequency was 2.8 SNP per kb in selected regions of rice and sorghum genomes (Feltus et al. 2004), 1 SNP per 70 bp in maize (Ching et al. 2002), and 1 SNP in 78, 189, and 9 bp in a particular gene(s) of grape (*Vitis vinifera*), barley (*Hordeum vulgare*), and wheat (*Triticum aestivum*), respectively (Kanazin et al. 2002; Caldwell et al. 2004; Salmaso et al. 2004). The nucleotide variation generated from six *EXPANSIN A* genes in interspecifiec cotton lines was similar to the SNP frequency of selected genes in other crops. However, Rong et al. (2004) showed that the nucleotide variation of intraspecific DNA sequence (*G. hirsutum* and *G. barbadense*) within amplicons derived from genetically mapped sequence-tagged sites (STS) was 0.35%.

Six *EXPANSIN A* gene sequences showed greater nucleotide variation in introns than in exons. Due to the large and direct effect on the phenotype, exons exhibit the least amount of nucleotide variation compared to introns (Holland et al. 2001). Polymorphisms occur at a frequency two to three times higher in introns than in exons in *Arabidopsis* and soybean, respectively (The Arabidopsis Genome Initiative 2000; Zhu et al. 2003). In cotton species, Wilkins et al. (1994) showed that intron sequences were less conserved than exons of the V-ATPase catalytic subunit superfamily, which was recently supported by comparative sequences analysis (Chee et al. 2004; Kumar et al. 2006).

Unequal evolutionary rate of the At- and Dt-genome in cotton

Results suggested a different evolutionary pattern of the six *EXPANSIN A* genes in the two subgenomes, such that the At-genome preferentially exhibits point mutations resulting in amino acid substitutions, while the Dt-genome preferentially displays short fragment deletions or insertions. The higher rate of SNP occurrence and nonsynonymous change in coding regions of the Dt-genome, and the nucleotide diversity comparison demonstrated that At- and Dt-genome of the tetraploid cotton evolve at unequal rates. These results were also supported by previous studies on phylogenetic analysis of low-copy gene sequences and molecular maker-based QTLs analysis. Phylogenetic analysis of *Adh* (Small et al. 1998) and *FAD2-1* (Liu et al. 2001) showed a faster evolutionary rate in the Dt-genome than in At-genome of tetraploid cotton. Furthermore, a study on the molecular evolutionary rate variation of 48 nuclear genes in cotton showed that D-Dt comparisons displayed higher

sequence divergence than in A-At comparisons (Senchina et al. 2003). Reinisch et al. (1994) reported that Dt-genome RFLP marker polymorphism levels were 10% higher than those of the At-genome. Moreover, inferences of the location of QTLs such as fiber-related traits (Jiang et al. 1998; Lacape et al. 2005), disease resistance (Wright et al. 1998), and leaf morphology (Jiang et al. 2000) repeatedly implied a higher evolutionary rate in the Dt-genome than in the At-genome of upland cotton. However, independent evolution of duplicated low- and single-copy sequences after polyploid formation in cotton has also been found in several independent phylogenetic analyses (Cronn et al. 1999; Small and Wendel 2000; Cedroni et al. 2003). These observations collectively indicate that the evolutionary forces on the two genomes may have been fundamentally different, and result in independent evolution of the homoeologous loci in tetraploid cotton.

Chromosomal locations of six *EXPANSIN A* genes

We utilized interspecific SNP markers by single nucleotide extension technology and deletion analysis (Liu et al. 2000) for chromosomal assignment of six *EXPANSIN A* genes. In cotton, the development of candidate gene markers based on conventional PCR methods has been limited by the paucity of genomic sequence data and relatively high levels of monomorphism for PCR amplicon mobility (Chee et al. 2004; Kumar et al. 2006). SNP markers derived from candidate genes would facilitate not only chromosomal assignment of these genes, but also QTL mapping and discovery of the roles of candidate genes in complex traits. In this study, we did not detect the chromosomal locations of duplicated homoeologous copies of *GhEXPA1*, *GhEXPA3*, *GhEXPA4*, *GhEXPA5*, and *GhEXPA6* genes. This could be due to the incomplete genome coverage in cytogenetic stocks, lacking homoeologous duplicated sequences, and relatively low interspecific nucleotide diversity. Our results revealed that all six *EXPANSIN A* genes may share a polyploid duplication event like *GhEXPA2* genes on the long arm of two homoeologous chromosomes 10 and 20 in tetraploid cotton. However, segmental duplication, an independent duplication event after polyploidization, might also play a role in gene evolution in cotton. For example, *GhEXPA2* and *GhEXPA3* both localize to the long arm of chromosome 10. Pfeil et al. (2004) suggested that the fate of the duplicated loci could be one of the several forms: (1) long-term maintenance of the same or similar function; (2) divergence in function; (3) loss of one copy; or (4) intralocus or interlocus gene conversion. Our results, showing chromosomal locations of some *EXPANSIN A* genes in non-homoeologous chromosomes, support earlier studies that polyploid nature in cotton created some unique avenues for response to selection (Wright et al. 1998; Rong et al. 2004). However for *rpb2* gene evolution in *Gossypium* species Pfeil et al. (2004) reported that single gene or segmental duplication was more likely the cause than ancient polyploidy. Rong et al. (2004) observed several duplication events within each subgenome, in addition to homoeologus duplication in cotton. They suggested that this could be due to retrotransposition, or present-day cotton may be derived from a putative ancestor containing six or seven chromosomes. Segmental duplications, as a part of polyploidization events, account for 12 out of 21 *EXPANSIN* genes in *Arabidopsis* and 16 out of 44 in rice (Sampedro et al. 2005). Identification of chromosomal locations of SNP markers using deletion analysis further confirmed the true allelic nature of the SNP markers and supported the merit of the strategy for discovery of SNP based on the phylogenetic and comparative analysis of sequences from tetraploid and closely related ancestral diploid species sequences for other candidate genes in allotetraploid cotton.

Fiber length is critical to fiber quality in the global cotton market. Identifying the genetic variants, diagnostic markers, and chromosomal locations of candidate genes that are associated with fiber cell elongation is likely to accelerate cotton improvement. Identification of SNPs for six *EXPANSIN A* genes may also help in functional genomics analysis because any changes in the nucleotides of the coding or regulatory regions of the gene may also have functional consequences. Paterson et al. (2003) reported one fiber length-related QTL on chromosome 20 with an LOD threshold of 3.75. They also found another fiber length QTL, significant in water-limited treatment, located on chromosome 9. We located one SNP marker specific to *GhEXPA1* and *GhEXPA2*, respectively, on the long arm of chromosome 20 and four different SNP markers derived from *GhEXPA4* on the long arm of chromosome 9 (Table 5). A QTL associated with fiber elongation was detected on chromosome 9 (Mei et al. 2004). Another QTL for fiber elongation was mapped on chromosome 3 (Ulloa et al. 2005), the same chromosome where we assigned four SNP markers generated from different regions of gene *GhEXPA6*. Furthermore, QTLs affecting fiber elongation were found on chromosome 1 and 20 where *GhEXPA1*, *GhEXPA2*, and *GhEXPA5* genes were also localized (Chee et al. 2005). Lacape et al. (2005) reported the association of fiber length QTLs with chromosome 3 and 10, and fiber elongation with chromosome 9, 10, and 20. QTL analysis by newly developed microsatellite markers suggested that loci on chromosome 3, 9, and 10 affect the fiber elongation, and 2.5 and 50% fiber span length, respectively (Frelichowski et al. 2006). The coincidence of chromosomal locations of the six *EXPANSIN A* genes and fiber length and elongation QTLs leaves open the possibility that the *EXPANSIN A* genes affect the important QTLs, especially considering their functional role in fiber cell expansion and elongation

(Smart et al. 1998; Arpat et al. 2004). SNP makers derived from candidate genes may be useful in exploring the roles of candidate genes in the complex traits.

In conclusion, we characterized *EXPANSIN* transcriptome expression profiles during cotton fiber polar elongation, developed an SNP marker discovery strategy in tetraploid cotton, characterized SNPs in the six cotton *EXPANSIN A* genes, and localized each of them to the chromosome. The overlap of the *EXPANSIN A* genes' chromosomal locations and previously reported fiber quality QTLs for fiber length and elongation may reflect the putative roles of some *EXPANSIN A* genes in these QTLs. These markers will have potential as candidate gene markers in marker-assisted selection program of fiber traits.

Acknowledgments We gratefully acknowledge the technical help of Mr. Douglas A. Dollar and the suggestions from Dr. Z. T. Burive and Ms. Yufang Guo. We also thank Dr. John Yu for providing *G. arboreum* and *G. raimondii* DNA samples.

References

Abdalla AM, Reddy OUK, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons using AFLP. Theor Appl Genet 102:222–229

Alvarez I, Cronn R, Wendel JF (2005) Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. Plant Syst Evol 252:199–214

Arpat AB, Waugh M, Sullivan JP, Gonzales M, Frisch D, Main D, Wood T, Leslie A, Wing RA, Wilkins TA (2004) Functional genomics of cell elongation in developing cotton fibers. Plant Mol Biol 54:911–929

Batley J, Barker G, O'Sullivan H, Edwards KJ, Edwards D (2003) Mining for single nucleotide polymorphisms and insertions/deletions in maize expressed sequence tag data. Plant Physiol 132:84–91

Brookes AJ (1999) The essence of SNPs. Gene 234:177–186

Brown MS, Menzel MY, Hasenkampf CA, Naqi S (1981) Chromosome configurations and orientations in 58 heterozygous translocations in *Gossypium hirsutum*. J Hered 72:161–168

Brubaker CL, Paterson AH, Wendel JF (1999) Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184–203

Bundock PC, Christopher JT, Eggler P, Ablett G, Henry RJ, Holton TA (2003) Single nucleotide polymorphisms in cytochrome P450 genes from barley. Theor Appl Genet 106:676–682

Bundock PC, Henry RJ (2004) Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. Theor Appl Genet 109:543–551

Caderas D, Muster M, Vogler H, Mandel T, Rose JKC, McQueen-Mason S, Kuhlemeier C (2000) Limited correlation between *EXPANSIN* gene expression and elongation growth rate. Plant Physiol 123:1399–1413

Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, Powell W (2004) Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. Genetics 167:941–947

Cedroni ML, Cronn RC, Adams KL, Wilkins TA, Wendel JF (2003) Evolution and expression of *MYB* genes in diploid and polyploid cotton. Plant Mol Biol 51:313–325

Chee PW, Draye X, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005) Molecular dissection of interspecific variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: I. Fiber elongation. Theor Appl Genet 111:757–763

Chee PW, Rong J, Williams-Coplin D, Schulze SR, Paterson AH (2004) EST derived PCR-based markers for functional gene homologues in cotton. Genome 47:449–462

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat Genet 23:203–207

Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. Theor Appl Genet 104:482–489

Cronn RC, Small RL, Wendel JF (1999) Duplicated genes evolve independently after polyploid formation in cotton. Proc Natl Acad Sci USA 96:14406–14411

Darley CP, Forrester AM, McQueen-Mason SJ (2001) The molecular basis of plant cell wall expansion. Plant Mol Biol 47:179–195

Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies: India and Japonica genome alignments. Genome Res 14:1812–1819

Frelichowski JE, Palmer MB, Main D, Tomkins JP, Cantrell RG, Stelly DM, Yu J, Kohel RJ, Ulloa M (2006) Cotton genome mapping with new microsatellites from Acala 'Maxxa' BAC-ends. Mol Genet Genomics 275:479–491

Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. Science 298:2345–2349

Glonek G, Solomon PJ (2004) Factorial and time course designs for cDNA microarray experiments. Biostatistics 5:89–111

Grover CE, Kim H, Wing RA, Paterson AH, Wendel JF (2004) Incongruent patterns of local and global genome size evolution in cotton. Genome Res 14:1474–1482

Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006) Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSR in allotetraploid cotton. Theor Appl Genet 112:430–439

Han ZG, Guo WZ, Song XL, Zhang TZ (2004) Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol Genet Genomics 272:308–327

Harmer SE, Orford SJ, Timmis JN (2002) Characterisation of six α-*expansin* genes in *Gossypium hirsutum* (upland cotton). Mol Genet Genomics 268:1–9

Holland JB, Helland SJ, Sharopova N, Rhyne DC (2001) Polymorphism of PCR-based markers targeting exons, introns, promoter regions, and SSRs in maize and intron and repeat sequences in oat. Genome 44:1065–1076

Jenkins JN, McCarty JC, Saha S, Gutierrez O, Hayes R, Stelly DM (2006) Genetic effects of thirteen *Gossypium barbadense* L. chromosome substitution lines in topcrosses with Upland cotton cultivars: I. Yield and yield components. Crop Sci 46:1169–1178

Ji SJ, Lu YC, Feng JX, Wei G, Li J, Shi YH, Fu Q, Liu D, Luo JC, Zhu YX (2003) Isolation and analysis of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. Nucleic Acids Res 31:2534–2543

Jiang C, Wright RJ, El-Zik K, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). Proc Natl Acad Sci USA 95:4419–4424

Jiang C, Wright RJ, Woo SS, DelMonte TA, Paterson AH (2000) QTL analysis of leaf morphology in tetraploid *Gossypium* (cotton). Theor Appl Genet 100:409–418

Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T (2002) Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*). Plant Mol Biol 48:529–537

Kende H, Bradford KJ, Brummell DA, Cho H-T, Cosgrove DJ, Fleming AJ, Gehring C, Lee Y, McQueen-Mason SJ, Rose JKC, Voesenek LACJ (2004) Nomenclature for members of the expansin superfamily of genes and proteins. Plant Mol Biol 55:1105–1110

Kerr MK, Churchill GA (2001) Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. Proc Natl Acad Sci USA 98:8961–8965

Kim MY, Van K, Lestart P, Moon JK, Lee SH (2005) SNP identification and SNAP marker development for a *GmNARK* gene controlling supernodulation in soybean. Theor Appl Genet 110:1003–1010

Kohel RJ, Yu J, Park YH, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. Euphytica 121:163–172

König R, Baldessari D, Pollet N, Niehrs C, Eils R (2004) Reliability of gene expression ratios for cDNA microarrays in multiconditional experiments with a reference design. Nucleic Acids Res 32:e29

Kumar P, Paterson AH, Chee PW (2006) Predicting intron sites by aligning cotton ESTs with *Arabidopsis* genomic DNA. J Cotton Sci 10:29–38

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinformatics 5:150–163

Kwok PY, Deng Q, Zakeri H, Nickerson DA (1996) Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. Genomics 31:123–126

Lacape JM, Nguyen TB, Courtois B, Belot JL, Giband M, Gourlot JP, Gawryziak G, Roques S, Hau B (2005) QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* × *Gossypium barbadense* backcross generations. Crop Sci 45:123–140

Li Y, Darley CP, Ongaro V, Fleming A, Schipper O, Baldauf SL, McQueen-Mason SJ (2002) Plant expansins are a complex multigene family with an ancient evolutionary origin. Plant Physiol 128:854–864

Liu Q, Brubaker CL, Green AG, Marshall DR, Sharp PJ, Singh SP (2001) Evolution of the *FAD2-1* fatty acid desaturase 5′ UTR intron and the molecular systematics of *Gossypium* (Malvaceae). Am J Bot 88:92–102

Liu S, Saha S, Stelly DM, Burr B, Cantrell RG (2000) Chromosomal assignment of microsatellite loci in cotton. J Hered 91:326–332

Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291

Menzel MY, Richmond KL, Dougherty BJ (1985) A chromosome translocation breakpoint map of the *Gossypium hirsutum* genome. J Hered 76:406–414

Mochida K, Yamazaki Y, Ogihara Y (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. Mol Genet Genomics 270:371–377

Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. Trends Plant Sci 9:325–330

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 19:233–240

Orford SJ, Timmis JN (1998) Specific expressioin of an *expansin* gene during elongation of cotton fibers. BBA—Gene Struct Expr 1398:342–346

Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellite in an interspecific recombinant inbred line cotton population. Mol Genet Genomics 274:428–441

Paterson AH, Saranga Y, Menz M, Jiang C, Wright RJ (2003) QTL analysis of genotype × environment interactions affecting cotton fiber quality. Theor Appl Genet 106:384–396

Pfeil BE, Brubaker CL, Craven LA, Crisp MD (2004) Paralogy and orthology in the Malvaceae *rpb2* gene family: investigation of gene duplication in *Hibiscus*. Mol Biol Evol 21:1428–1437

Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847

Richert AM, Premstaller A, Gebhardt C, Oefner PJ (2002) Genotyping of SNPs in a polyploid genome by pyrosequencing. Biotechniques 32:592–593,596–598,600

Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding X, Garza JJ, Marler BS, Park C, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Pateson AH (2004) A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496–2497

Ruan YL, Llewellyn DJ, Furbank RT (2001) The control of single-celled cotton fiber elongation by developmentally reversible gating of plasmodesmata and coordinated expression of sucrose and K^+ transporters and expansion. Plant Cell 13:47–60

Saha S, Jenkins JN, Wu J, McCarty JC, Gutierrez OA, Percy RG, Cantrell RG, Stelly DM (2006a) Effects of chromosome-specific introgression in Upland cotton on fiber and agronomic traits. Genetics 172:1927–1938

Saha S, Raska DA, Stelly DM (2006b) Upland cotton (*Gossypium hirsutum* L.) × Hawaiian cotton (*G. tomentosum* Nutt. ex Seem.) F_1 hybrid hypoaneuploid chromosome substitution series. J Cotton Sci 12:263–272

Salmaso M, Faes G, Segala C, Stefanini M, Salakhutdinov I, Zyprian E, Toepfer R, Grando MS, Velasco R (2004) Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L.), as revealed by single nucleotide polymorphisms. Mol Breeding 14:385–395

Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6:242

Sampedro J, Lee Y, Carey RE, dePanphilis C, Cosgrove DJ (2005) Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. Plant J 44:409–419

Senchina DS, Alwarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol Biol Evol 20:633–643

Shappley ZW, Jenkins JN, Meredith WR, McCarty JC (1998) An RFLP linkage map of Upland cotton, *Gossypium hirsutum* L. Theor Appl Genet 97:756–761

Shimizu Y, Aotsuka S, Hasegawa O, Kawada T, Sakuno TF, Hayashi T (1997) Changes in the level of mRNAs for cell wall-related enzymes in growing cotton fiber cells. Plant Cell Physiol 38:375–378

Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Am J Bot 85:1301–1315

Small RL, Wendel JF (2000) Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). Genetics 155:1913–1925

Smart LB, Vojdani F, Maeshima M, Wilkins TA (1998) Genes involved in osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated. Plant Physiol 116:1539–1549

Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 49:431–437

Stelly DM, Saha S, Raska DA, Jenkins JN, McCarty JC, Gutierrez OA (2005) Registration of 17 Upland (*Gossypium hirsutum*) cotton germplasm lines disomic for different *G. barbadense* chromosome or arm substitutions. Crop Sci 45:2663–2665

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Ulloa M, Saha S, Jenkins JN, Meredith WR, McCarty JC, Stelly DM (2005) Chromosomal assignment of RFLP linkage groups harboring important QTLs on an intraspecific cotton (*Gossypium hirsutum* L.) joinmap. J Hered 96:132–144

Van K, Hwang EY, Kim MY, Park HJ, Lee SH, Cregan PB (2005) Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea. J Hered 96:529–535

Vogler H, Caderas D, Mandel T, Kuhlemeier C (2003) Domains of *expansin* gene expression define growth regions in the shoot apex of tomato. Plant Mol Biol 53:267–272

Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082

Wendel JF, Cronn C (2003) Polyploidy and the evolutionary history of cotton. Adv Agron 78:139–186

Wilkins TA, Arpat AB (2005) The cotton fiber transcriptome. Physiologia Plantarum 124:295–300

Wilkins TA, Jernstedt JA (1999) Molecular genetics of developing cotton fiber. In: Basra AS (ed) Cotton fiber: development biology, quality improvement, and textile processing. Binghamton, New York, pp 231–269

Wilkins TA, Smart LB (1996) Isolation of RNA from plant tissue. In: Krieg PA (ed) A laboratory guide to RNA: isolation, analysis, and synthesis. Wiley-Liss, New York, pp 21–41

Wilkins TA, Wan CY, Lu CC (1994) Ancient origin of the vacuolar H^+-ATPase 69-kilodalton catalytic subunit superfamily. Theor Appl Genet 89:514–524

Wright RJ, Thaxton PM, El-Zik KM, Paterson AH (1998) D-subgenome bias of *Xcm* resistance genes in tetraploid *Gossypium* (cotton) suggests that polyploid formation has created novel avenues for evolution. Genetics 149:1987–1996

Zhang J, Guo W, Zhang T (2002) Molecular linkage mapping of allotetraploid cotton (*Gossypium hirsutum* L. × *Gossypium barbadense* L.) with a haploid population. Theor Appl Genet 105:1166–1174

Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR (2003) Identification of SNPs and development of allele-specific PCR markers for γ-gliadin alleles in *Triticum aestivum*. Theor Appl Genet 107:130–138

Zhao X, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH (1998) Dispersed repetitive DNA has spread to new genome since polyploid formation in cotton. Genome Res 8:479–492

Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163:1123–1134